DOI 10.1002/pmic.200900459 *Proteomics* 2010, *10*, 634–649

REVIEW

De novo sequencing of peptides by MS/MS

Joerg Seidler, Nico Zinn, Martin E. Boehm and Wolf D. Lehmann

Molecular Structure Analysis, German Cancer Research Center, Heidelberg, Germany

The current status of *de novo* sequencing of peptides by MS/MS is reviewed with focus on collision cell MS/MS spectra. The relation between peptide structure and observed fragment ion series is discussed and the exhaustive extraction of sequence information from CID spectra of protonated peptide ions is described. The partial redundancy of the extracted sequence information and a high mass accuracy are recognized as key parameters for dependable *de novo* sequencing by MS. In addition, the benefits of special techniques enhancing the generation of long uninterrupted fragment ion series for *de novo* peptide sequencing are highlighted. Among these are terminal ^{18}O labeling, MS^n of sodiated peptide ions, N-terminal derivatization, the use of special proteases, and time-delayed fragmentation. The emerging electron transfer dissociation technique and the recent progress of MALDI techniques for intact protein sequencing are covered. Finally, the integration of bioinformatic tools into peptide *de novo* sequencing is demonstrated.

Received: June 30, 2009
Revised: September 18, 2009
Accepted: September 23, 2009

Keywords:
b ion series / CID / Manual sequencing / Sequence information / y ion series

1 Introduction

The introduction of chemical protein sequencing by Edman degradation [1] in the 1950s was a milestone in the development of protein research. For identification of Edman degradation products, mainly LC was used. Intermittently, MS was introduced as alternative method to LC with optical detection. This was achieved in combination with the early soft ionization techniques chemical ionization [2], field desorption [3], and fast atom bombardment [4]. Later, the gradual refinement of MS/MS techniques created the basis for peptide sequencing by MS (for a review, see [5]), which finally gave fast access to multiple internal protein sequences by the analysis of proteolytic peptides [6–8]. In a parallel development termed ladder sequencing, protein sequences were read by single stage MALDI-MS from sequence ladders generated by exopeptidases [9–12]. Within the last two decades, protein sequence determination by MS/MS became more and more powerful, a development driven mainly by the improvements of LC techniques in combination with ESI [13]. The advantages of MS/MS techniques with respect to speed, sensitivity, and applicability to complex peptide mixtures gradually led to the replacement of Edman techniques by LC-MS/MS. As a result, MS-based proteomics has emerged as the method of choice for the identification of proteins [14, 15] *via* database-supported interpretation of MS data using search engines such as MASCOT [16], SEQUEST [17], X! TANDEM [18], or OMSSA [19]. MS data are typically either MALDI mass fingerprint data [20, 21] or LC-ESI-MS/MS data [22, 23]. More recently, the introduction of MALDI-MS/MS has strengthened the role of MALDI for amino acid sequence determination. Considering all amino acid sequence combinations that are theoretically possible, only a very minor portion of protein sequences is realized in nature, and therefore a short peptide sequence is already highly protein-specific. This situation effects that a database-supported, probability-based annotation of peptide MS/MS spectra leads to protein identification at a high level of confidence from fragmentary sequence information.

Thus, database-supported protein identification is very effective, but it precludes the recognition of all peptides not

Correspondence: Professor Wolf D. Lehmann, Molecular Structure Analysis (W160), German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
E-mail: wolf.lehmann@dkfz.de
Fax: +49-6221-42-45-54

Abbreviations: ECD, electron capture dissociation; **ETD,** electron transfer dissociation; **ISD,** in-source decay; **sCID,** skimmer-CID; **SPITC,** 4-sulphophenyl isothiocyanate; **TDF,** time delayed fragmentation

present in the reference database. In spite of the continuously growing sequence databases, *de novo* sequencing of peptides, *i.e.* sequencing without assistance of a linear sequence database, is still essential in several analytical situations. For example, analyses of protein sequence variants or their splice isoforms require *de novo* sequencing, as well as protein analysis from organisms with unsequenced genomes. In addition, *de novo* sequencing is essential for analysis of peptides containing non-proteinic or modified amino acids, as typically present, *e.g.* in bioactive peptides of bacteria or fungi [24]. The performance of both the MS/MS and of the LC part influences the utility of an LC-MS system for *de novo* sequencing. This is because the significance of peptide MS/MS data is connected with the purity of the peptide ions selected for fragmentation. In the following, the state-of-the-art and current advancements of peptide *de novo* sequencing by MS are reviewed and discussed. Important points are the redundancy of sequence information present in CID spectra, the use of new peptide activation techniques, MS^n analyses, the influence of mass accuracy, labeling techniques, the fragmentation of intact proteins, and bioinformatic tools for a reference-free interpretation of MS/MS spectra.

2 Nomenclature of peptide fragment ions

The introduction of soft ionization techniques enabled the efficient generation of intact peptide ions which can be selected as precursors for subsequent activation and detection of their fragment ions. For peptide sequencing, the positive mode is generally used due to its higher sensitivity and since MS/MS spectra of protonated peptides contain a wealth of sequence-specific fragment ions. In contrast, negative ion MS/MS spectra contain less sequence information and are usually more difficult to interpret. Fragmentation of protonated peptide ions following CID occurs predominantly at the peptide backbone by proton-induced fragmentation reactions as explained by the mobile proton model [25]. The widely accepted nomenclature for the annotation of peptide sequence ions is outlined in Fig. 1, as originally proposed by Roepstorff and Fohlman [26] and later modified by Biemann [27]. Sequence ions that undergo neutral loss are distinguished further as for example the loss of H_2O is indicated by $^\circ$ or the loss of NH_3 by *.

3 CID

The majority of peptide *de novo* sequencing has been performed using CID (or collision-activated dissociation). In low energy CID, activation of molecular ions is achieved by collisions with inert gas molecules (He, N_2, Ar) present in a separate collision cell or as bath gas in an ion trap. In CID, the translational energy of the ions is partially converted into internal vibrational energy which then induces peptide

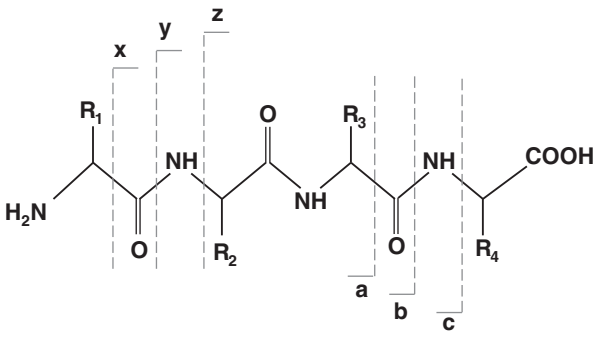

Figure 1. Nomenclature of sequence-specific peptide fragments; a-, b-, and c-type ions contain the N-terminus; x-, y-, and z- ions contain the C-terminus; hydrogen rearrangements are omitted in this simplified annotation (according to [27]). R1, R2, R3 represent the side chains of the amino acid residues.

fragmentation. Acceleration of the precursor ions in the collision zone by a potential difference (collision offset) intensifies the fragmentation and leads to more second (or higher order) fragments, due to repeated collisions during the travel through the collision cell. Moderate collision offset values are most favorable for sequencing, in terms of absolute fragment ion intensity and background level. For unmodified peptides, activation of protonated molecules leads primarily to backbone fragmentations, resulting in structure-specific fragments. The mobile proton model [25] can semi-quantitatively rationalize the distribution of the observed fragments [28, 29]. A particular feature of backbone cleavages of multiply protonated molecular ions is that they may result in complementary b/y fragment ions. Cleavages at the N-terminal site of P or at the C-terminal site of D are the prototype forms of fragmentation resulting in complementary fragment ions. However, in collision cell MS/MS spectra of peptides the occurrence of complementary ions is rather the exception than the rule. This implies that the majority of backbone cleavages are only represented by either b or y ions and that the individual C- or N-terminal fragment ion series mostly exhibit only a minor overlap. This situation is demonstrated in Fig. 2 for a collection of quadrupole TOF (Q-TOF) CID spectra summarizing their search engine-annotated fragment ion series. The data in Fig. 2 imply that the location of the basic residues R, K, and H determines the balance between the length of the b and y ion series.

The CID spectra of typical tryptic peptides with a basic residue at the C-terminus are characterized by extended y ion series accompanied by short b ion series (b_2 to b_3, b_4) without b_1 ions (Fig. 2A). Peptides with a basic residue at the N-terminus may be generated, *e.g.* by digestion with LysN. Their CID spectra are characterized by long b ion series in connection with short y ion series (Fig. 2B). Peptides with basic residues at both termini exhibit CID spectra with b and y ion series of comparable length often with moderate overlap (Fig. 2C), and peptides with tryptic

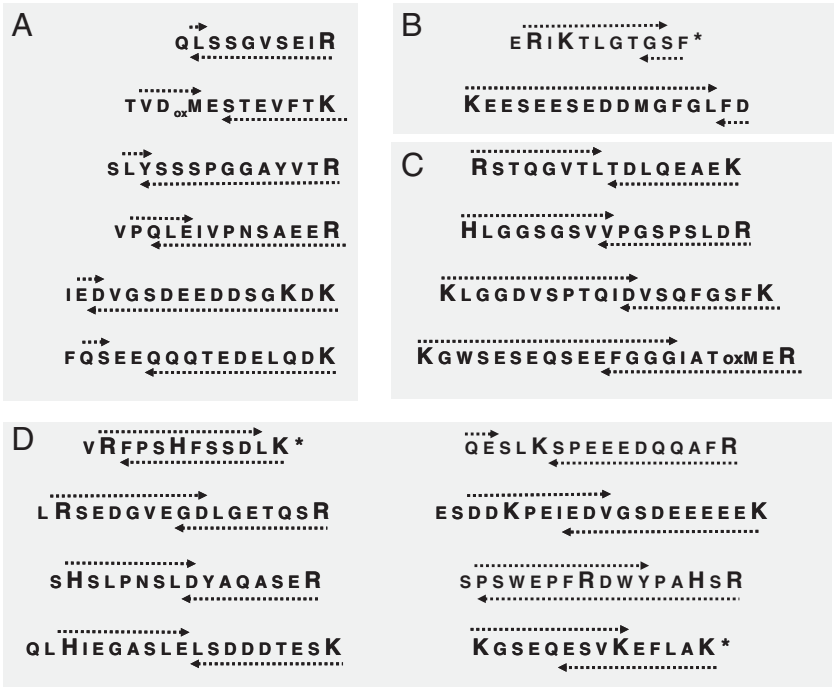

Figure 2. Distribution of b (→) and y (←) series fragment ions in Q-TOF CID spectra as observed in a set of peptides with different distributions of basic amino acids (data from [30], * data from [31]); (A) peptides with a basic residue at the C-terminus; (B) peptides with a basic residue at the N-terminus; (C) peptides with basic residues at both termini; (D) peptides with terminal and internal basic residues. (Explanation of the annotation for QLSSGVSEIR: the arrows indicate the presence of b_2 and of the y series from y_1 to y_9.)

miscleavage sites show a similar characteristic. In general, their occurrence is caused by either a KK, RR, KR, RK, KP, or RP motif or by the presence of an acidic residue in one of the sequence positions ± 3 relative to K or R [32, 33]. For these peptides with internal basic residues, a more pronounced mutual overlap of the fragment ion series or sequence ion suppression around internal basic residues may occur (Fig. 2D). In practice, backbone cleavage in the vicinity of R and K residues is not very effective due to sequestering of the protons at the basic site. Thus, the corresponding ions are of low relative abundance, so that they may escape detection, as observed for the peptide internal K starting with QES (Fig. 2D). The presence of two or more R residues in direct or close vicinity normally causes pronounced gaps in fragment ion series of any type, due to the high basicity of R. ESI peptide ions with charge state 2+ and 3+ are the preferred precursor ions for *de novo* sequencing. In practice, the MS/MS spectra of all accessible charge states should be recorded, since the corresponding spectra often exhibit b or y ion series of different length, so that different regions of the peptide may become accessible in this way. In general, fragment ion spectra of higher charge states contain more sequence ions; however, MS/MS spectra of doubly charged ions are often more easily interpreted than those of triply or quadruply charged precursor ions. The manual annotation of the basic ion series has been explicitly summarized in a step-by-step tutorial [34].

Besides b and y ions, additional types of fragment ions occur in CID spectra of peptides which also contain sequence information. These comprise neutral loss reactions from the peptide termini [35, 36] and internal fragment ions [37]. This is exemplified for the *de novo* sequencing of the peptide SNTDANQ[L/I]WT[L/I]K shown in Fig. 3. This peptide was obtained by tryptic digestion of a type II ribosome-inactivating protein from the plant *Ximenia americana*. Its sequence could be determined on the basis of the almost complete y ion series from y_1 to y_{max-2} (Fig. 3A). In addition, a continuous b ion series is observed from b_2 to b_8-NH_3 (b_8^*). Inspection of the low mass side of the $[M+2H]^{2+}$ signal revealed the occurrence of neutral loss fragmentations, which can be easily recognized by their +2 charge state (Fig. 3B). These ions show the N-terminal loss of S followed by the loss of N, which resolves the ambiguity in the structure of the b_2 ion at m/z 202 between SN and NS. The occurrence of the N-terminal sequence SN is also supported by the b_2 ion fragment profile. This profile shows the loss of NH_3 (a_2^* and b_2^*), which is typical for b_2 SN, whereas b_2 NS ions show the preferential loss of H_2O (reference data not shown). The assignment of sequence and compositional isomers of b_2 ions *via* their fragmentation profiles has been demonstrated recently [38, 39]. The low mass region of the MS/MS spectrum (Fig. 3C) also provides compositional information in the form of immonium ions.

Table 1 summarizes the sequence and compositional information, which can be extracted from the MS/MS spectrum in Fig. 3. It can be recognized that redundant information is obtained for the N-terminus based on the b_2 ion fragmentation profile and on the neutral losses from the molecular ion. The central sequence TDANQ is based both on the occurrence of b and y ions. Finally, the occurrence of the C-terminal K residues is supported both by the fragment ion mass and by the specificity of trypsin. The result is rounded by the detection of the immonium ions for Q, W, and [L/I].

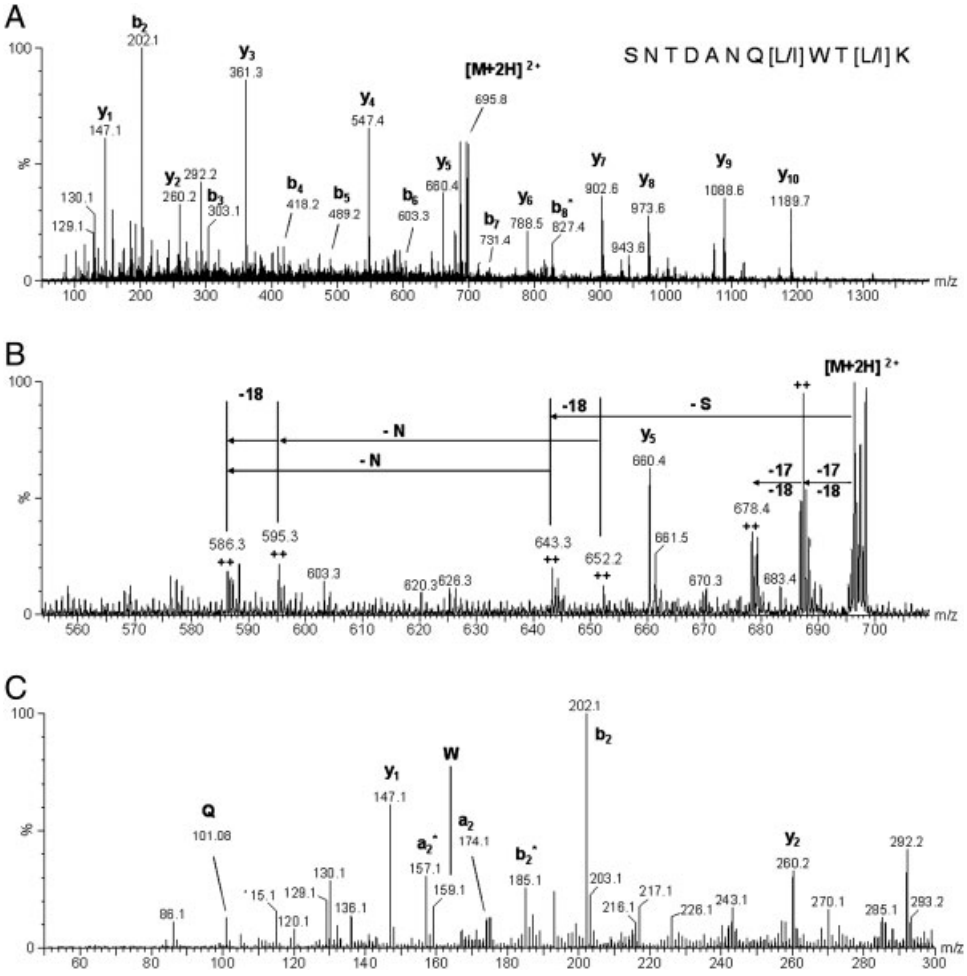

Figure 3. *De novo* sequencing by Q-TOF CID of the tryptic peptide SNTDANQ[L/I]WT[L/I]K originating from a type II ribosome- inactivating protein isolated from *Ximenia americana*; (A) complete MS/MS spectrum characterized by y and b ion series; (B) expanded central region showing C-terminal neutral losses; (C) expanded low mass region showing the y_1, y_2, and the b_2 ion and its fragments, as well as immonium ions.

Complementary b/y ion pairs do not only give sequence information but also provide additional molecular weight information, since the sum of their mass values is equivalent to the precursor ion mass. This is helpful in case a mixture of precursor ions or a precursor ion of very low abundance is isolated. In addition to the sequence information summarized in Table 1, peptide MS/MS spectra sometimes show internal b type fragment ions [37]. Finally, reporter fragmentations for covalent modifications are highly useful for recognition of modified peptides [40]. Over all, the occurrence of redundant sequence information contained in several types of fragmentations of multiply charged ions is a beneficial feature of ESI in combination with collision cell CID, which strongly favors its use for *de novo* sequencing of peptides. Collision cell MS/MS spectra are highly dependent on the collision offset used. The extent of fragmentation may be varied from only partial to complete decomposition of the molecular ion by increasing the offset value. Simultaneously, fragment ions undergo secondary (and higher order) fragmentations. Nevertheless, due to the large body of experimental MS/MS data available, optimal offset values can be selected empirically.

4 Comparison of collision cell and ion trap CID

Peptide fragment ion spectra generated in a collision cell or in an ion trap are both generated by CID. Nevertheless, they exhibit differences caused by the way of activation, mass analysis, and time scale. In principle, ion trap MS/MS spectra are very reproducible and dominated by first generation fragments, since fragments are normally not further activated due to their different *m/z* value. Thus, typically more complementary b/y ion pairs are observed compared to Q-TOF MS/MS spectra. This leads to the occurrence of pronounced overlaps between b and y ion series, even for normal tryptic peptides with a single basic residue at the C-terminus (Fig. 4). This characteristic creates redundant sequence information, which puts the read-out of the sequence on a reliable basis.

Table 1. Types of sequence information present in the MS/MS spectrum in Fig. 3

	S	N	T	D	A	N	Q	[L/I]	W	T	[L/I]	K
y Ions			*	*	*	*	*	*	*	*	*	*
b Ions		*	*	*	*	*	*	*				
Neutral loss	*	*										
b_2 Fragment profile	*	*										
Composition							*	*	*		*	
Protease specificity												*

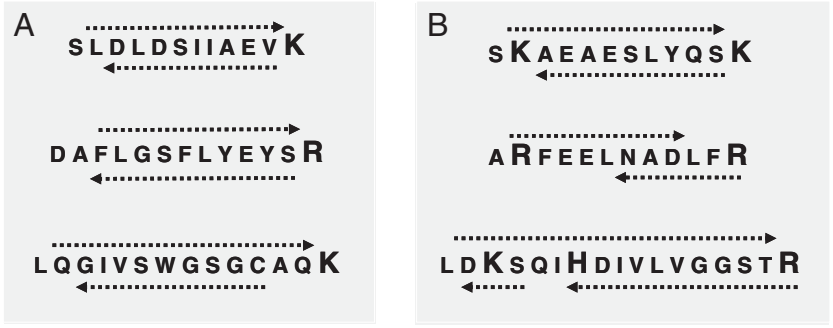

Figure 4. Typical fragment ion series distribution in peptides observed in linear ion trap MS/MS spectra; (A) tryptic peptides with a single basic site; (B) tryptic peptides with internal basic residues. For both groups of peptides, b and y ion series with pronounced overlap are observed.

For *de novo* sequencing, ion trap MS/MS spectra have a certain limitation of a ''low mass cut-off,'' which is directly proportional to the m/z value of the precursor ion. The average size of tryptic peptides excludes fragment ion detection below about m/z 200–300, limiting the information about the peptide ends present in low mass fragments. Using stepwise MS^n, low mass fragment ions can also be detected (see below). Recently developed operating conditions for ion traps (pulsed Q-dissociation) allow a generally enhanced detection of low mass ions [41]. Hybrid instruments such as the LTQ-orbitrap combination offer an additional activation mode occurring in the transfer region between the two mass analyzers, which generates MS/MS spectra with highly abundant low mass fragments [42].

Currently, MS^n in an ion trap appears to be the most applicable technique for differentiation between leucine and isoleucine. In the early days of MS/MS, high-energy CID was more frequently applied than today. Using this ion activation method for peptide analysis, side chain fragmentations were observed. These were found to be useful for the differentiation between leucine and isoleucine. Peptide fragments containing L were found to show satellites at −42 Da, compared to −28 Da satellites for I [43]. In low-energy CID of peptides containing L/I, the immonium ion of L/I at m/z 86 is frequently observed with high abundance. Further MS/MS analysis of this immonium ion showed that the fragment ion at m/z 69 is specific for I [44–46]. Ion traps are particularly favorable for such an analysis [41] due to their MS^n capability. Such an analysis is shown in Fig. 5 providing clear evidence that the residue in the peptide GpSVAVGVIK at the position 2 from the C-terminus is isoleucine.

5 MALDI-PSD, MALDI-TOF/TOF, and MALDI in-source decay (MALDI-ISD)

MALDI-PSD was introduced as the first MALDI technique for detection of fragment ions [47]. Fragments formed by metastable fragmentation between the ion source and the reflector can be detected in this way. Subsequently the uninterrupted detection of a complete MALDI-PSD spectrum was demonstrated [48], in contrast to the original stitching of several partial PSD spectra. MALDI-PSD spectra of peptides show b and y ion series and neutral losses as observed following CID. However, since MALDI generates mainly singly charged molecular ion, MALDI-PSD spectra contain less sequence information compared to the CID spectra of multiply charged precursor ions as generated by ESI. The fragmentation of MALDI generated peptide ions was put on an improved instrumental basis by the development of two types of MALDI-TOF/TOF instruments [49, 50] allowing the recording of peptide MS/MS spectra with improved sequence information [51, 52]. In addition, several attempts were undertaken to improve the sequence information of MALDI-TOF/TOF spectra of peptides by derivatization [53]. Most frequently, the derivatization is targeted to the peptide N-terminus with the aim to enhance the C-terminal fragment ion series (y ions) which are relatively stable species allowing simplified sequencing [54]. Furthermore the pK values of the α-amino group at the N-terminus and the ε-amino group of lysine are often sufficiently apart allowing their selective derivatization. In contrast, this is not feasible for the C-terminus, since the pK values between C-terminal and side chain carboxy functions are less clearly separated. Amino group specific modification can be

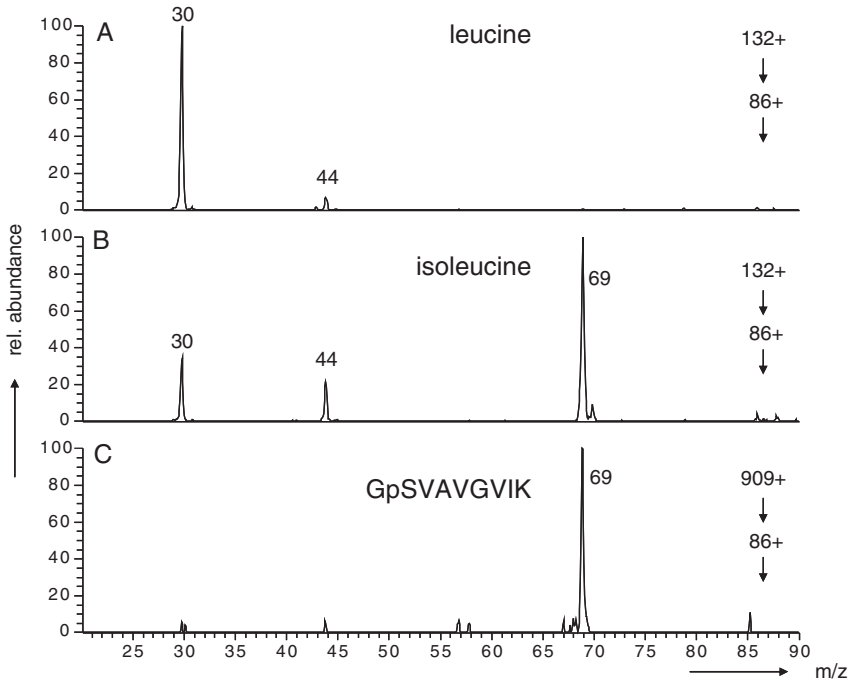

Figure 5. Ion trap MS/MS spectra of *m/z* 86 for differentiation between leucine and isoleucine; (A) MS/MS of 86 generated from leucine; (B) MS/MS of *m/z* 86 generated from isoleucine; (C) MS/MS of *m/z* 86 generated from the peptide GpSVAVGVIK. The high abundance of *m/z* 69 in (c) identifies the presence of isoleucine.

achieved with sulfonic acid derivatives [55]. For arginine-containing tryptic peptides this leads to a specific N-terminal modification. Lysine-containing peptides can be modified equally following guanidylation of lysine ε-amino groups [56]. Currently, Lys guanidylation is often combined with derivatization by 4-sulphophenyl isothiocyanate (SPITC) [57–59] and 2-sulfobenzoic acid anhydride [56, 60]. An example for the improvement of a MALDI fragment ion spectrum is given in Fig. 6, adapted from a recent *de novo* sequencing study of plant protein isoforms [61]. The MALDI fragment ion spectrum of the peptide YVTYAA[I/L]AGDASV[I/L]DDR shows the same y fragment ions before and after N-terminal derivatization with SPITC. However, after derivatization b ions are practically absent and the y ions are of high and nearly uniform abundance. The suppression of b ions by SPITC derivatization is explained by the replacement of the N-terminal amino function by the sulfonic acid function, so that the derivatized b fragments are not able to stabilize an extra proton.

An approach for influencing the fragmentation behavior of peptides in MALDI-PSD refers to tryptic peptides with a C-terminal lysine, which is reacted with a strongly basic reagent [62, 63], resulting in the exclusive occurrence of a y ion series.

By the introduction of delayed extraction MALDI [64], the detection of fragment ions generated in the MALDI source became feasible (ISD) [65]. MALDI-ISD detects fragment ions formed in the time gap between the laser pulse and the switch to the full accelerating voltage (100–500 ns). The typical MALDI-ISD fragment ions of peptides are of *c* and *z* type, pointing toward a high abundance of radicals in the MALDI plume. The same type of fragments are observed in ExD fragmentation techniques (see below), where they originate from precursor radical ions generated by electron transfer. Matrix optimization for MALDI-ISD has been described [66]. Applications for database-supported protein identification using the top down approach have been reported, *e.g.* [67, 68] and the technique appears to offer potential also for *de novo* sequencing. The usefulness of MALDI for *de novo* sequencing is currently further strengthened by the availability of MALDI-MS/MS instruments (*e.g.* MALDI-QTOF, MALDI-TOF/TOF, MALDI-Orbitap).

6 ExD techniques

Recently, electron capture dissociation (ECD) [69] and electron transfer dissociation (ETD) [70] (summarized as ExD techniques) have been introduced as new activation techniques for peptide fragmentation, which can be regarded as complementary to CID [71]. ExD techniques function *via* the transfer of a single low energy electron to a multiply protonated peptide, mostly with charge state 3+ or 4+, as often observed in ESI. The electron is transferred either directly (ECD) or from a previously formed radical anion (ETD). The free radical site introduced upon electron transfer leads to an instantaneous and local radical-induced backbone cleavage. This cleavage affects mainly the N-C bond, so that the resulting fragment ion spectra typically show c or z type ions. A detailed mechanism for their formation has been proposed [70]. Recently, the combination of metalloendopeptidase LysN digestion with ETD or MALDI-TOF/TOF has been applied for *de novo* peptide

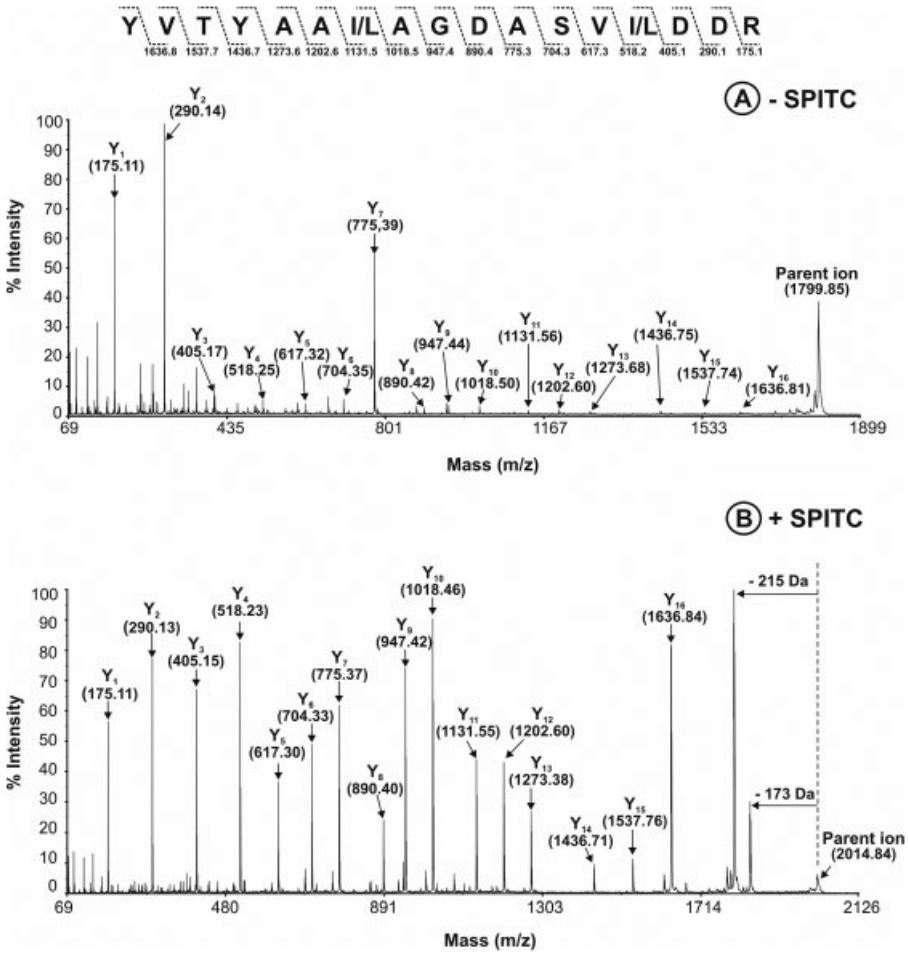

Figure 6. MALDI MS/MS spectra of $[M+H]^+$ions of the peptide YVTYAA[I/L]AGDASV[I/L]DDR (A) before and (B) after SPITC derivatization. Derivatization induces neutral losses of the modifying group (−173 Da, −215 Da) and a pronounced enhancement of all y ions. Adapted from [61], with permission.

sequencing [72, 73]. Peptides generated by LysN carry a K residue at the N-terminus resulting in the preferential occurrence of c ion series, a situation facilitating a read-out of the peptide sequence.

ExD spectra are well suited for sequencing of modified peptides [74], since fragmentations originating from side chains are normally not observed. Even relatively labile structures such as phosphorylated amino acids stay intact upon ExD. However, variations in the peptide backbone structure influence the radical induced fragmentation. For instance, cleavage at the N-terminal site of P is suppressed due to the existence of two N-C bonds, an observation that is in accordance with the proposed fragmentation mechanism. Another example is the presence of an isoaspartyl residue, introducing an extra C-C bond into the peptide/protein backbone, which is cleaved using ECD [75, 76] or ETD [77]. This cleavage leads to isoAsp-specifc backbone fragments. For isoAsp at position n from the N-terminus, a $c_{n-1}+58$ ion is formed, and a complementary $z_{m-1}+57$ may be formed, where m annotates the isoAsp position counted from the C-terminus. A relative quantification of the site-specific isoAsp content in peptides has been demonstrated by ECD [78]. Such an analysis has also been demonstrated using CID [79]; however, ExD techniques appear to be superior, since they present isoAsp-specific ions, whereas CID techniques mainly display quantitative changes in fragment ion abundances [79] or isoAsp-specific ions of minor relative abundance [80].

7 LC elution behavior and *de novo* Sequencing

In LC-MS/MS analyses, the peptide retention time is an analytical parameter, which is obtained without extra effort. Nowadays, the LC retention times of peptides can be predicted with high reliability (in a window of $\pm$2–3 min) for different chromatographic systems, *e.g.* [81–83]. The relative elution order is also a useful parameter, as can be demonstrated in the analysis of the just mentioned isoAsp-containing peptides. Since they are generated during protein aging, they are generally accompanied by their cognates carrying a normal aspartyl residue. On RP-LC, peptides with internal isoAsp peptides elute before their unmodified analogs, whereas peptides with N-terminal isoaspartate elute later [84] (peptides with C-terminal isoas-

partate do not exist). In this way, the LC retention time may be useful as an additional ''soft'' parameter to confirm the result of a *de novo* sequencing step.

8 Mass accuracy and peptide sequencing

The reliability of peptide sequencing improves with increasing accuracy of the mass measurement, since the exact mass contains information about the elemental composition. The most obvious benefit of accurate mass measurement refers to the correct assignment of fragment ion series. K and Q are nominally isobaric amino acids with a mass difference of about 36 mDa, and the difference between W and EG is about 15 mDa. These ambiguities can be differentiated at medium resolution, as provided by Q-TOF instruments. However, interfering ions of different types with a multitude of mass differences including very small values may occur, so that high mass accuracy is of general value.

Accurate mass data are without use for the recognition of structural isomers since these are characterized by identical elemental compositions. For differentiation between isoelemental ions, differences in their fragmentation behavior or in their chemical properties, *e.g.* in derivatization or label exchange reactions have to be employed. First of all, the pair L/I cannot be differentiated by mass measurement. In addition, numerous accurate mass numbers exist, which represent different amino acid combinations. For several dipeptide combinations, a single elemental composition may be connected with four structures (*e.g.* AN, NA, GQ, and QG). Differentiation of these quartets present as N-terminal motifs can be achieved *via* the b_2 ion fragmentation profile [39]. Other examples of isoelemental structures are the pairs GG/N (both $C_4H_6N_2O_2$) and GA/Q (both $C_5H_8N_2O_2$).

Nowadays several instrument types (*e.g.* FT-ICR, OrbiTrap) are available [85], providing MS analysis with a resolution in the range of 100.000–500.000 with mass accuracies in the range of 2–0.2 ppm. An early example for the new possibilities offered by such extreme mass accuracy is the introduction of composition-based *de novo* sequencing [86, 87]. This two-step procedure starts with calculation of a set of possible amino acid compositions on the basis of the highly accurate mass value for a peptide molecular ion and its fragment ions. Based on the calculation of all mass value-compatible amino acid compositions, a database containing all permutations is generated. This database is then used for assignment of the MS/MS spectra. The new possibilities of highly accurate mass data for automated *de novo* sequencing have been summarized recently [88].

9 *De novo* sequencing and stable isotope labeling

Stable isotope labeling techniques can also be used to facilitate *de novo* sequencing. Selective labeling of the C-terminus by incorporation of ^{18}O is particularly useful for this purpose, since it is a relatively fast and cost-effective method. The first attempts for ^{18}O introduction were performed by acid-catalyzed exchange and by esterase-catalyzed cleavage of methyl esters [89] with subsequent analysis by fast atom bombardment. However, acid treatment leads to pronounced peptide hydrolysis and esterase cleavage is relatively inefficient due to a low affinity of the esterase toward peptides. In contrast, tryptic digestion in $H_2^{18}O$ effects a fast incorporation of one or two ^{18}O atoms at the C-terminus of tryptic peptides, as visualized by ESI and MALDI-MS [90]. As expected, a selective labeling of y ions was obtained in this way. Digestion in a mixture of $H_2^{16}O$ and $H_2^{18}O$ generates y ions with a characteristically distorted isotope pattern [91–93] enabling their straightforward differentiation from the unlabeled b ions. The ^{18}O label can also be introduced after digestion in a separate step [94]. The combination of tryptic digestion with ^{18}O labeling appears to be particularly useful, since tryptic peptides tend to show long uninterrupted y ion series, and since by principle, internal y ions do not occur. As an example, Fig. 7 shows the MS/MS spectrum of the peptide STDANQ[L/I]WT[L/I]K, partially labeled with ^{18}O at the carboxy terminus. All y ions can be recognized by an elevated +2 Da signal in their isotopic envelopes.

Differential terminal derivatizations with a labeled and a nonlabeled reagent have been introduced both for quantitative proteomics as well as for facilitated *de novo* sequencing. The latter purpose is achieved, since N-terminal labeling shifts only the b ion series, whereas C-terminal labeling selectively affects the corresponding y ions. This concept has been demonstrated for derivatization with H_4/D_4-(Nicotinoyloxy)succinimide [95], O-methylisourea [96], D_4-lysine [97], H_2/D_2-formaldehyde [98] or more recently by N-terminal H_3/D_3-acetylation using acetic acid anhydride [99].

10 Special MS/MS techniques suited for *de novo* peptide sequencing

10.1 Multistage fragmentation of sodium-cationized peptides

Clear-cut C-terminal sequence information from peptides can be obtained by multiple stage MS [100, 101] in an ion trap. This is particularly the case, when lithium- or sodium-cationized peptides instead of protonated peptides are selected as precursor ions. The prevailing fragmentation process of $[M+Li]^+$ and $[M+Na]^+$ ions is the neutral loss of the C-terminal amino acid building block with formation of an $[M+Cat]^+$ ion of the same type, but shortened by one amino acid. Thus, the process can be repeated using the MS^n capabilities of an ion trap, as demonstrated in several investigations. A sequence ladder generated in this way is shown in Fig. 8 for the peptide SQGIASTK.

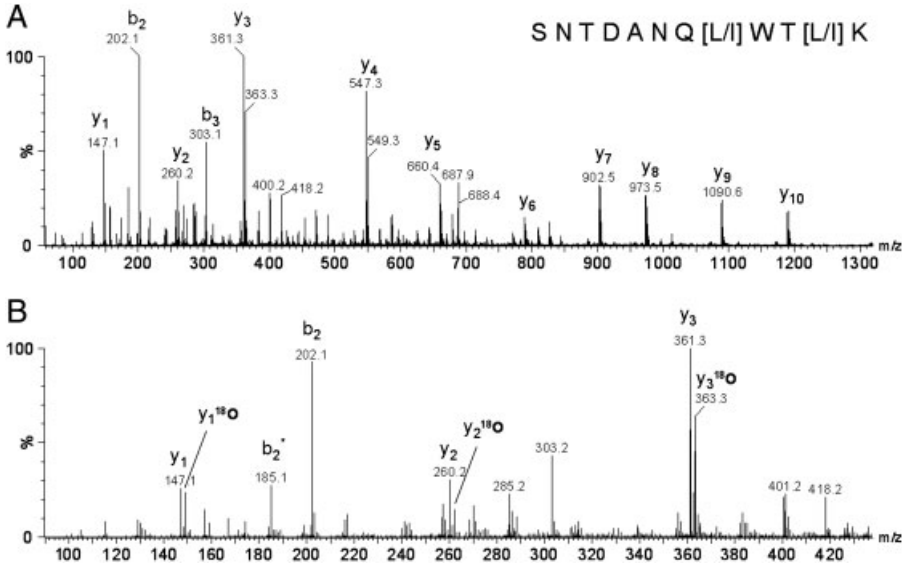

Figure 7. *De novo* sequencing by Q-TOF CID of the tryptic peptide STDANQ[L/I]WT[L/I]K, partially labeled with ^{18}O at its carboxy terminus; (A) complete spectrum; (B) expanded low mass region, demonstrating the facile differentiation between unlabeled b ions and partially ^{18}O-labeled y ions.

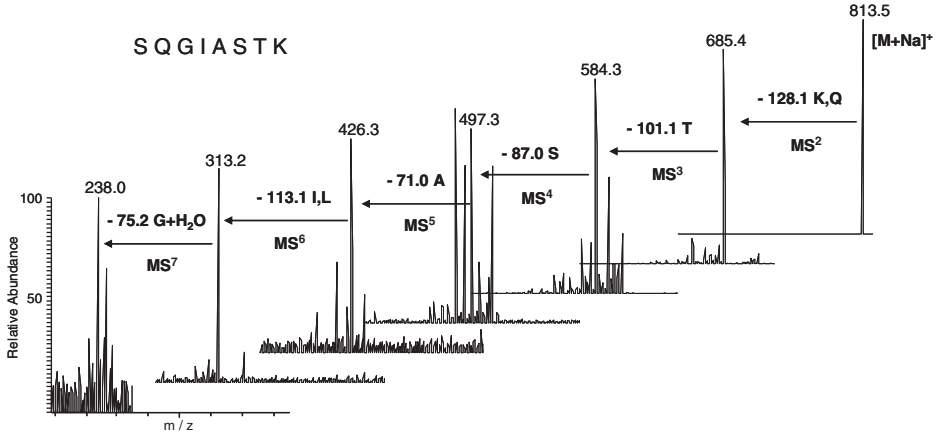

Figure 8. Ion trap multistage MS/MS for C-terminal sequencing of the peptide SQGIASTK as $[M+Na]^+$ ion. Shown is the overlay of six spectra, generated by repeated fragmentation of the stepwise shortened sodiated peptide ion.

Unfortunately, the broader application of this elegant C-terminal sequencing method is limited by the fact that no methods are available for the preferred generation of cationized peptides at high sensitivity.

10.2 Time-delayed fragmentation

The distribution of product ions observed following CID is time-dependent. Using a Q-Trap instrument, selective recording of MS/MS spectra containing preferentially the more stable y ions (compared to b ions) has been demonstrated. The term ''time delayed fragmentation'' (TDF) [102, 103] has been created to describe the detection of fragment ions produced at variable time windows and thus subpopulations of ions with different internal energy. Three major steps are involved in TDF: (i) ion activation, (ii) ion relaxation, and (iii) fragment collection. TDF provides the ability to simplify the product ion spectra as shown in Fig. 9. In case all fragment ions produced by CID are collected, the MS/MS spectrum in Fig. 9A is obtained, showing both b and y ions. Relaxation of the precursor ions in the linear ion trap over 10 ms leads to a population of precursor ions with lower internal energy. Selective recording of fragment ions originating from this relaxed subpopulation leads to a simplified MS/MS spectrum containing exclusively y ions as shown in Fig. 9B. In this way, TDF supports *de novo* sequencing of peptides.

10.3 Single series fragment ion spectra

Simplified fragment ion spectra of peptides which exhibit only C-terminal or N-terminal fragment ions can be recorded by using a two stage fragmentation. Instrumentally, this can be realized using a triple quadrupole analyzer with

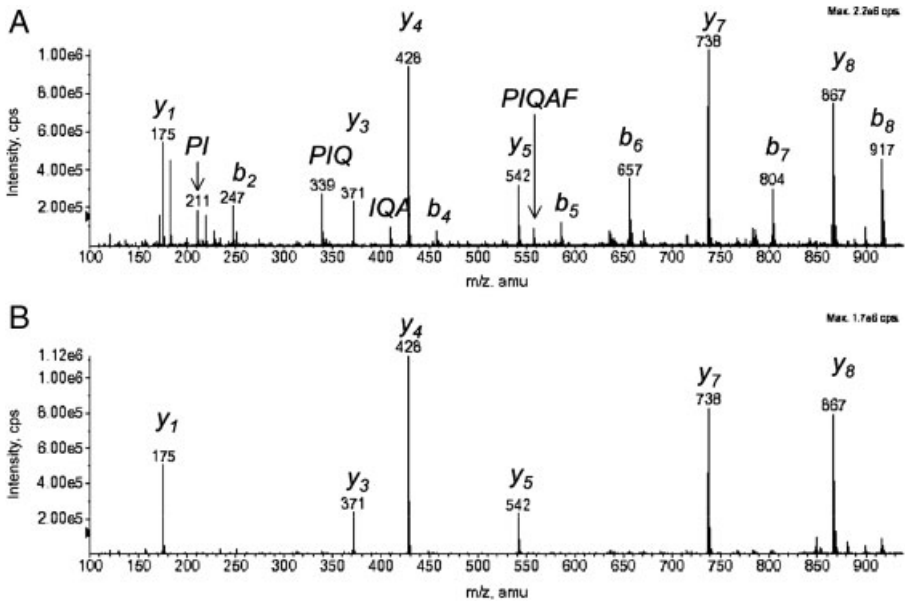

Figure 9. Q-trap MS/MS spectra of the doubly charged β-casein fragment at *m/z* 1094 (DMPIQAFLLYQEPVLPGPVR) using (A) conventional CID with fragmentation window up to about 250 μs, and (B) TDF with a fragmentation window > 10 ms. Adapted from [104], with permission.

combined skimmer-CID (sCID) and collision cell CID. This is a pseudo-MS^3 analysis, since skimmer CID represents a fragmentation mode without precursor ion selection. Thus, this experimental set up will only provide clear results for pure samples containing a single peptide. Nevertheless, two-stage CID for peptide sequencing has particular merits (Fig. 10). This is demonstrated for the T1 fragment of protein kinase A, which is a short N-terminally myristoylated heptapeptide. A selective detection of b or y ions is achieved using a b ion or y ion-specific precursor ion scan, respectively. A common b ion fragment is found at *m/z* 211, which represents the myristoyl fragment, and a common y ion fragment is the y_1 ion. Therefore, using skimmer CID in combination with precursor ion scanning for *m/z* 211, only b ions are detected, whereas the combination with precursor ion scanning for 147 leads to pure y ion spectra.

A general applicability of two-stage CID for the recording of single series peptide MS/MS spectra as demonstrated in Fig. 10 requires the combination of two genuine MS/MS steps, as can be realized, *e.g.* in a penta-quadrupole instrument with two collision cells. However, this type of instrument is commercially not available.

11 Automated *de novo* sequencing

Engines for *de novo* sequencing have undergone continuous improvement [106–112] as tools for database-supported spectra interpretation. Basic or full versions of several of these tools are available in the internet, as mentioned in the corresponding publications. Currently, all results provided by automated *de novo* sequencing should be checked manually. Nevertheless, we found automated *de novo* sequencing very helpful for extracting meaningful MS/MS spectra from large sets of MS/MS spectra, thus increasing the success rate of manual *de novo* sequencing. Long sequence ion series from central parts of peptides were in general correctly recognized, as well as complementary b/y ion pairs. Inconsistencies between proposed and manual sequences were often found near the terminal regions, where characteristic details in the spectra were not fully recognized. As an example, Table 2 shows three automated annotations from a dataset acquired for *de novo* sequencing of a type II ribosome-inactivating protein from *X. americana*, in comparison with manual sequencing results and the later determined sequence of riproximin (Q2PA54).

The results in Table 2 show that the annotation tool may interchange a single amino acid with an isobaric two amino acid combination. This error is caused by the false-positive recognition of a sequence ion. Other errors refer to the annotation of the peptide ends. Since the CID spectra in general contain several independent sources of information for terminal sequences (Table 1), this situation shows that the further refinement of *de novo* annotation tools toward a better recognition of the terminal sequence is a promising route for their improvement.

Proteins of organisms with unknown genome often show sequence homologies to functionally related homologs in other already completely sequenced organisms. Therefore, existing sequence databases may be helpful as support for *de novo* sequencing results obtained from unknown proteins. In practice, peptide sequence candidates may be subjected to homology-based search programs such as BLAST (basic local alignment search tool) or the FASTA (fast all) algorithm to identify similarities to sequences present in databases [113–115]. An extensive review on homology-driven proteomics has been given recently [116]. Bioinformatic tools are of high interest for further advancement of *de novo* peptide sequencing.

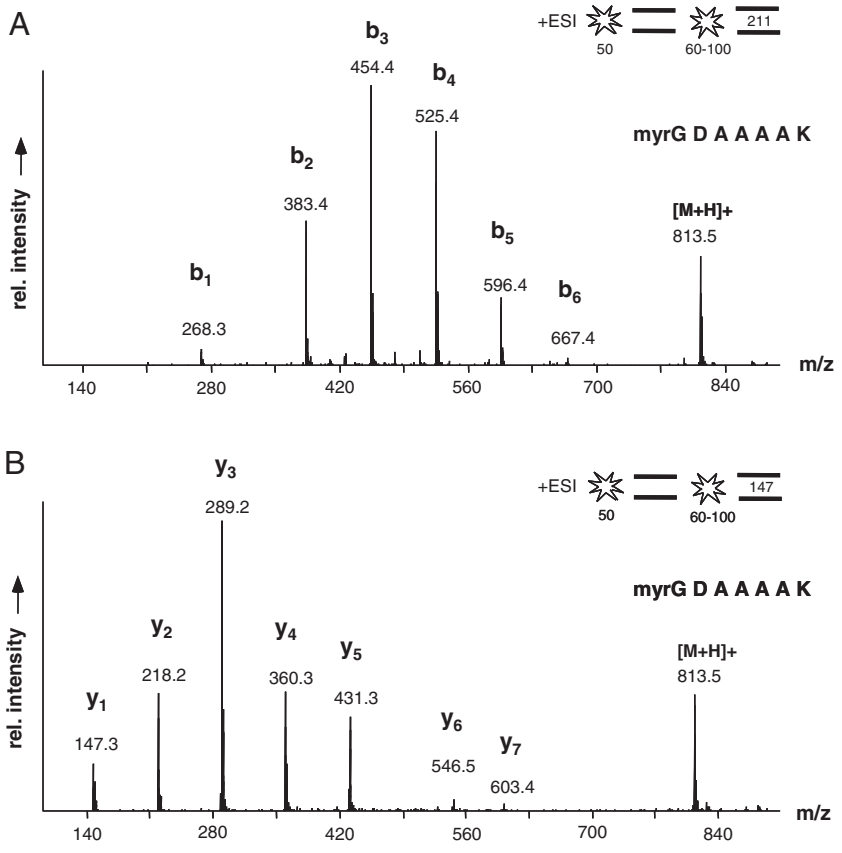

Figure 10. Triple quadrupole pseudo MS^3 spectra of the peptide myrGDAAAAK derived from protein kinase A; (A) sCID+ precursor scan for 211 (myristoyl fragment ion); (B) sCID+precursor ion scan for 147 (y_1 ion of K). These scan modes selectively shows only the b ion series A) or the y ion series b), respectively (see also Ref. [105]).

Table 2. Automated and manual *de novo* sequencing (http:// www.bioinfor.com/peaksonline) of peptides extracted from *Ximenia americana*

Precursor ion	Sequences	Type of sequence
[687.3] 2+	N S A D A N G A [L/I] E G N [L/I] K	Proposed
	S N A D A N Q [L/I] W N [L/I] K	Manual
[703.9] 2+	T E Q Q W A [L/I] Y P S S P	Proposed
	T E Q Q W A [L/I] Y P D R	Manual
	A E Q K W A L F P D R	Riproximin (505–515)
[695.8] 2+	S N T D A N G A [L/I] E G T [L/I] K	Proposed
	S N T D A N Q [L/I] W T [L/I] K	Manual
	S N T D A N Q L W I L K	Riproximin (374–385)

Some deviations between automated and manual sequencing are observed. Two peptide sequences were found to be homologous to riproximin (Q2PA54).

12 Complete protein sequencing supported by *de novo* MS/MS peptide sequencing

Complete protein sequence may be obtained by MS alone by the implementation of different proteases. Each protease generates different sets of peptides so that overlapping sequences may be found and longer continuous sequences can be obtained. In general, a sequence overlap of three amino acids is sufficient for stitching of two sequences, since mostly a three amino acid sequence motif is already unique for a protein of intermediate size (20–60 kDa). The combined use of trypsin, chymotrypsin and AspN is beneficial for this purpose, due to their different cleavage characteristics at basic, neutral, or acidic sites. In elastase digests numerous overlapping peptides are found, due to its low specificity. In spite of this lack of specificity, elastase digestions leads to peptides of around 1 kDa size, since the affinity of elastase to its substrates strongly decreases when the peptide length falls below 0.8–1 kDa. Peptides of this length can often be sequenced completely by CID, unless multiple basic sites are present. The first complete primary

structure determination of an entire protein by MS alone was performed using multiprotease digestion and MS/MS [117]. To our knowledge, the largest protein that has been entirely sequenced using the bottom-up approach so far is a 21-kDa cytochrome c_4 [118]. In connection with the bottom-up approach, two phenomena may cause errors in the primary sequence. First, rearrangement processes during ion trap CID have been described [119]; second, protease-catalyzed transpeptidation reactions have been observed, leading, *e.g.* to the transfer of terminal residues or to the ligation of originally distant sequence parts [120, 121]. Although these phenomena have been clearly described, their actual impact on the confidence of *de novo* sequencing results is not yet explored.

The introduction of ExD techniques has improved the technical basis for ''top-down'' protein sequencing, *i.e.* the fragmentation of complete proteins. For example, a sequence coverage of about 70% on the first ~200 residues was demonstrated for each terminus of large proteins grater than 200 kDa [122]. A top-down approach for targeted characterization of C- and N-termini of undigested proteins based on MALDI-ISD has also been introduced under the synonym ''T^3-sequencing'' [67]. Yoo and co-workers recently demonstrated the potency of this technique by sequencing a 31 residue polyethylene glycol modified peptide completely [123]. Concerning the addressed challenges and benefits of both – ''bottom-up'' and ''top-down'' – approaches it is most likely that they will continue to co-evolve in future or will meet halfway as hybrid approaches, in which large fragments or whole domains of proteins are analyzed intact [124]. Peptide *de novo* sequencing results can also help to obtain protein sequences *via* molecular biology approaches. For instance, a cDNA library is constructed and primers derived from *de novo* sequenced peptides are used for RACE-PCR (rapid amplification of cDNA-ends with polymerase chain reaction) [125, 126]. Finally the complete protein sequence is identified using DNA sequencing.

13 Concluding remarks

CID and ETD are currently the most effective techniques employed in peptide *de novo* sequencing. Both techniques generate peptide MS/MS spectra with very high structural information. In this context, CID is best suited for small peptides of 1–2 kDa, whereas ETD can cope with larger peptides. In the quest for large precursor ions, MALDI-ISD has shown remarkable progress. The results of all ionization and fragmentation techniques benefit from the increased mass accuracy of MS/MS data, reducing the number of sequences compatible with a MS/MS dataset. In case of ambiguities in the *de novo* sequencing of peptides, several chemical and instrumental methods exist for improving the specificity of the results. However, due to their extra efforts, they will probably be applied in selective cases only. Complete peptide *de novo* sequencing by MS/MS will not be generally successful due to interfering factors, such as (i) a low intensity causing incomplete detection of sequence ions, (ii) a peptide sequence preventing the formation of a sufficient set of sequence ions, or (iii) the presence of an unusual amino acid and/or an uncommon covalent modification. Nevertheless, once a complete peptide sequence can be read from an MS/MS spectrum, this result has a high level of confidence. This is particularly the case when the information redundancy present in the majority of peptide MS/MS spectra is used and when additional data, such as a protease specificity, LC elution behavior. or fragmentation rules, are integrated into the final evaluation. The further refinement of automated *de novo* sequencing tools in relation to high resolution MS/MS data may complement the widely applied database-supported search algorithms. In this way, not database-filed protein sequence variations, which now remain unassigned in an automated annotation, could also be reliably identified. Thus, in many proteomic analyses the combination of database-supported annotation with automated *de novo* sequencing will probably further advance the interpretation of MS/MS data.

The authors have declared no conflict of interest.

14 References

[1] Edman, P., Method for the determination of the amino acid sequence in peptides. *Acta Chem. Scand.* 1950, *4*, 283–293.

[2] Fales, H. M., Nagai, Y., Milne, G. W. A., Brewer, H. B. *et al.*, Chemical ionization mass spectrometry of complex molecules 7 use of chemical ionization mass spectrometry in analysis of amino acid phenylthiohydantoin derivatives formed during Edman degradation of proteins. *Anal. Biochem.* 1971, *43*, 288–299.

[3] Schulten, H. R., Wittmann-Liebold, B., High-resolution field desorption mass-spectrometry 5. Mixtures of amino-acid phenylthiohydantoins and Edman degradation products. *Anal. Biochem.* 1976, *76*, 300–310.

[4] Bradley, C. V., Williams, D. H., Peptide sequencing using the combination of Edman degradation, carboxypeptidase digestion and fast atom bombardment mass-spectrometry. *Biochem. Biophys. Res. Commun.* 1982, *104*, 1223–1230.

[5] Biemann, K., Laying the groundwork for proteomics – Mass spectrometry from 1958 to 1988. *Int. J. Mass Spectrom.* 2007, *259*, 1–7.

[6] Biemann, K., Martin, S. A., Mass-spectrometric determination of the amino-acid-sequence of peptides and proteins. *Mass Spectrom. Rev.* 1987, *6*, 1–75.

[7] Hunt, D. F., Bone, W. M., Shabanowitz, J., Rhodes, J., Ballard, J. M., Sequence-analysis of oligopeptides by secondary ion-collision activated dissociation mass-spectrometry. *Anal. Chem.* 1981, *53*, 1704–1706.

[8] Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S., Hauer, C. R., Protein sequencing by tandem mass-spectrometry. *Proc. Natl. Acad. Sci. USA* 1986, *83*, 6233–6237.

[9] Chait, B. T., Wang, R., Beavis, R. C., Kent, S. B. H., Protein ladder sequencing. *Science* 1993, *262*, 89–92.

[10] Cool, D. R., Hardiman, A., C-Terminal sequencing of peptide hormones using carboxypeptidase Y and SELDI-TOF mass spectrometry. *Biotechniques* 2004, *36*, 32–34.

[11] Patterson, D. H., Tarr, G. E., Regnier, F. E., Martin, S. A., C-terminal ladder sequencing via matrix-assisted laser-desorption mass-spectrometry coupled with carboxypeptidase-Y time-dependent and concentration-dependent digestions. *Anal. Chem.* 1995, *67*, 3971–3978.

[12] Thiede, B., Wittmann-Liebold, B., Bienert, M., Krause, E., MALDI-MS for C-terminal sequence determination of peptides and proteins degraded by carboxypeptidase-Y and carboxypeptidase-P. *Febs Lett.* 1995, *357*, 65–69.

[13] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray ionization for mass-spectrometry of large biomolecules. *Science* 1989, *246*, 64–71.

[14] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, *422*, 198–207.

[15] Cox, J., Mann, M., Is proteomics the new genomics? *Cell* 2007, *130*, 395–398.

[16] Perkins, D. N., Pappin, D. J. C., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551–3567.

[17] Yates, J. R., Eng, J. K., Clauser, K. R., Burlingame, A. L., Search of sequence databases with uninterpreted high-energy collision-induced dissociation spectra of peptides. *J. Am. Soc. Mass Spectrom.* 1996, *7*, 1089–1098.

[18] Craig, R., Beavis, R. C., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004, *20*, 1466–1467.

[19] Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L. *et al.*, Open mass spectrometry search algorithm. *J. Proteome Res.* 2004, *3*, 958–964.

[20] Gevaert, K., Vandekerckhove, J., Protein identification methods in proteomics. *Electrophoresis* 2000, *21*, 1145–1154.

[21] Pappin, D. J. C., Hojrup, P., Bleasby, A. J., Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* 1993, *3*, 327–332.

[22] Domon, B., Aebersold, R., Review – mass spectrometry and protein analysis. *Science* 2006, *312*, 212–217.

[23] Mann, M., Wilm, M., Error tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* 1994, *66*, 4390–4399.

[24] Degenkolb, T., Brückner, H., Peptaibiomics: towards a myriad of bioactive peptides containing C-alpha-dialkyl-amino acids? *Chem. Biodiv.* 2008, *5*, 1817–1843.

[25] Wysocki, V. H., Tsaprailis, G., Smith, L. L., Breci, L. A., Special feature: commentary – mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* 2000, *35*, 1399–1406.

[26] Roepstorff, P., Fohlman, J., Proposal for a common nomenclature for sequence ions in mass-spectra of peptides. *Biomed. Mass Spectrom.* 1984, *11*, 601–601.

[27] Biemann, K., Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol.* 1990, *193*, 886–887.

[28] Zhang, Z. Q., Prediction of low-energy collision-induced dissociation spectra of peptides. *Anal. Chem.* 2004, *76*, 3908–3922.

[29] Steen, H., Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* 2004, *5*, 699–711.

[30] Imanishi, S. Y., Kochin, V., Ferraris, S. E., de Thonel, A. *et al.*, Reference-facilitated phosphoproteomics – fast and reliable phosphopeptide validation by microLC-ESI-Q-TOF MS/MS. *Mol. Cell. Proteomics* 2007, *6*, 1380–1391.

[31] Seidler, J., Adal, M., Kübler, D., Bossemeyer, D., Lehmann, W. D., Analysis of autophosphorylation sites in the recombinant catalytic subunit alpha of cAMP-dependent kinase by nanoUPLC-ESI-MS/MS. *Anal. Bioanal. Chem.* 2009, *395*, 1713–1720.

[32] Thiede, B., Lamer, S., Mattow, J., Siejak, F. *et al.*, Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion. *Rapid Commun. Mass. Spectrom.* 2000, *14*, 496–502.

[33] Winter, D., Kugelstadt, D., Seidler, J., Kappes, B., Lehmann, W. D., Protein phosphorylation influences proteolytic cleavage and kinase substrate properties exemplified by analysis of in vitro phosphorylated PfGAP45 by nanoUPLC-MS/MS. *Anal. Biochem.* 2009, *393*, 41–47.

[34] Kinter, M., Sherman, N. E., *Protein Sequencing and Identification Using Tandem Mass Spectrometry*, Wiley Interscience, New York 2000.

[35] Martin, D. B., Eng, J. K., Nesvizhskii, A. I., Gemmill, A., Aebersold, R., Investigation of neutral loss during collision-induced dissociation of peptide ions. *Anal. Chem.* 2005, *77*, 4870–4882.

[36] Salek, M., Lehmann, W. D., Neutral loss of amino acid residues from protonated peptides in collision-induced dissociation generates N- or C-terminal sequence ladders. *J. Mass Spectrom.* 2003, *38*, 1143–1149.

[37] Schlosser, A., Lehmann, W. D., Patchwork peptide sequencing: extraction of sequence information from accurate mass data of peptide tandem mass spectra recorded at high resolution. *Proteomics* 2002, *2*, 524–533.

[38] Winter, D., Lehmann, W. D., Individual b(2) ion fragmentation profiles combined with AspN digestion improve N-terminal peptide sequencing. *Anal. Bioanal. Chem.* 2009, *393*, 1587–1591.

[39] Winter, D., Lehmann, W. D., Sequencing of the thirteen structurally isomeric quartets of N-terminal dipeptide motifs in peptides by collision-induced dissociation. *Proteomics* 2009, *9*, 2076–2084.

[40] Hung, C. W., Schlosser, A., Wei, J. H., Lehmann, W. D., Collision-induced reporter fragmentations for identification of covalently modified peptides. *Anal. Bioanal. Chem.* 2007, *389*, 1003–1016.

[41] Cunningham, C., Glish, G. L., Burinsky, D. J., High amplitude short time excitation: A method to form and detect low mass product ions in a quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* 2006, *17*, 81–84.

[42] Olsen, J. V., Macek, B., Lange, O., Makarov, A. *et al.*, Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* 2007, *4*, 709–712.

[43] Johnson, R. S., Martin, S. A., Biemann, K., Collision-induced fragmentation of [M+H]+ions of peptides – side-chain specific sequence ions. *Int. J. Mass Spectrom. Ion Processes* 1988, *86*, 137–154.

[44] Armirotti, A., Millo, E., Damonte, G., How to discriminate between leucine and isoleucine by low energy ESI-TRAP MS^n. *J. Am. Soc. Mass Spectrom.* 2007, *18*, 57–63.

[45] Hulst, A. G., Kientz, C. E., Differentiation between the isomeric amino acids leucine and isoleucine using low-energy collision-induced dissociation tandem mass spectrometry. *J. Mass Spectrom.* 1996, *31*, 1188–1190.

[46] Nakamura, T., Nagaki, H., Ohki, Y., Kinoshita, T., Differentiation of leucine and isoleucine residues in peptides by consecutive reaction mass-spectrometry. *Anal. Chem.* 1990, *62*, 311–313.

[47] Kaufmann, R., Spengler, B., Lutzenkirchen, F., Mass-spectrometric sequencing of linear peptides by product-ion analysis in a reflectron time-of-flight mass-spectrometer using matrix-assisted laser-desorption. *Rapid Commun. Mass. Spectrom.* 1993, *7*, 902–910.

[48] Cornish, T. J., Cotter, R. J., A curved field reflectron time-of-flight mass-spectrometer for the simultaneous focusing of metastable product ions. *Rapid Commun. Mass. Spectrom.* 1994, *8*, 781–785.

[49] Medzihradszky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M. *et al.*, The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* 2000, *72*, 552–558.

[50] Suckau, D., Resemann, A., Schuerenberg, M., Hufnagel, P. *et al.*, A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal. Bioanal. Chem.* 2003, *376*, 952–965.

[51] Shui, W. Q., Liu, Y. K., Fan, H. Z., Bao, H. M. *et al.*, Enhancing TOF/TOF-based de novo sequencing capability for high throughput protein identification with amino acid-coded mass tagging. *J. Proteome Res.* 2005, *4*, 83–90.

[52] Samyn, B., Sergeant, K., Memmi, S., Debyser, G. *et al.*, MALDI-TOF/TOF de novo sequence analysis of 2-D PAGE-separated proteins from Halorhodospira halophila, a bacterium with unsequenced genome. *Electrophoresis* 2006, *27*, 2702–2711.

[53] Roth, K. D. W., Huang, Z. H., Sadagopan, N., Watson, J. T., Charge derivatization of peptides for analysis by mass spectrometry. *Mass Spectrom. Rev.* 1998, *17*, 255–274.

[54] Barofsky, D. F., Chen, T. F., Peptide sequencing promoted by N-terminal derivatization. *Mass Spectrom. Hyphen. Tech. Neuropeptide Res.* 2002, 345–373.

[55] Keough, T., Youngquist, R. S., Lacey, M. P., Sulfonic acid derivatives for peptide sequencing. *Anal. Chem.* 2003, *75*, 156A–165A.

[56] Samyn, B., Debyser, G., Sergeant, K., Devreese, B., Van Beeumen, J., A case study of de novo sequence analysis of N-sulfonated peptides by MALDI TOF/TOF mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2004, *15*, 1838–1852.

[57] Marekov, L. N., Steinert, P. M., Charge derivatization by 4-sulfophenyl isothiocyanate enhances peptide sequencing by post-source decay matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* 2003, *38*, 373–377.

[58] Chen, P., Nie, S., Mi, W., Wang, X. C., Liang, S. P., De novo sequencing of tryptic peptides sulfonated by 4-sulfophenyl isothiocyanate for unambiguous protein identification using post-source decay matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass. Spectrom.* 2004, *18*, 191–198.

[59] Wang, D. X., Kalb, S. R., Cotter, R. J., Improved procedures for N-terminal sulfonation of peptides for matrix-assisted laser desorption/ionization post-source decay peptide sequencing. *Rapid Commun. Mass. Spectrom.* 2004, *18*, 96–102.

[60] Keough, T., Youngquist, R. S., Lacey, M. P., A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry. *Proc. Natl. Acad. Sci. USA* 1999, *96*, 7131–7136.

[61] Rinalducci, S., Roepstorff, P., Zolla, L., De novo sequence analysis and intact mass measurements for characterization of phycocyanin subunit isoforms from the blue-green alga Aphanizomenon flos-aquae. *J. Mass Spectrom.* 2009, *44*, 503–515.

[62] Conrotto, P., Hellman, U., Lys Tag: an easy and robust chemical modification for improved de novo sequencing with a matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometer. *Rapid Commun. Mass. Spectrom.* 2008, *22*, 1823–1833.

[63] Peters, E. C., Horn, D. M., Tully, D. C., Brock, A., A novel multifunctional labeling reagent for enhanced protein characterization with mass spectrometry. *Rapid Commun. Mass. Spectrom.* 2001, *15*, 2387–2392.

[64] Brown, R. S., Lennon, J. J., Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser-desorption ionization linear time-of-flight mass-spectrometer. *Anal. Chem.* 1995, *67*, 1998–2003.

[65] Brown, R. S., Lennon, J. J., Sequence-specific fragmentation of matrix-assisted laser-desorbed protein peptide ions. *Anal. Chem.* 1995, *67*, 3990–3999.

[66] Demeure, K., Quinton, L., Gabelica, V., De Pauw, E., Rational selection of the optimum MALDI matrix for top-down proteomics by in-source decay. *Anal. Chem.* 2007, *79*, 8678–8685.

[67] Suckau, D., Resemann, A., T-3-sequencing: targeted characterization of the N- and C-termini of undigested proteins by mass spectrometry. *Anal. Chem.* 2003, *75*, 5817–5824.

[68] Hardouin, J., Protein sequence information by matrix-assisted laser desorption/ionization in-source decay mass spectrometry. *Mass Spectrom. Rev.* 2007, *26*, 672–682.

[69] Horn, D. M., Zubarev, R. A., McLafferty, F. W., Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc. Natl. Acad. Sci. USA* 2000, *97*, 10313–10317.

[70] Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. USA* 2004, *101*, 9528–9533.

[71] Zubarev, R. A., Zubarev, A. R., Savitski, M. M., Electron capture/transfer versus collisionally activated/induced dissociations: Solo or duet? *J. Am. Soc. Mass Spectrom.* 2008, *19*, 753–761.

[72] Boersema, P. J., Taouatas, N., Altelaar, A. F. M., Gouw, J. W. *et al.*, Straightforward and de novo peptide sequencing by MALDI-MS/MS using a Lys-N metalloendopeptidase. *Mol. Cell. Proteomics* 2009, *8*, 650–660.

[73] Taouatas, N., Drugan, M. M., Heck, A. J. R., Mohammed, S., Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat. Methods* 2008, *5*, 405–407.

[74] Chi, A., Huttenhower, C., Geer, L. Y., Coon, J. J. *et al.*, Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc. Natl. Acad. Sci. USA* 2007, *104*, 2193–2198.

[75] Cournoyer, J. J., Pittman, J. L., Ivleva, V. B., Fallows, E. *et al.*, Deamidation: Differentiation of aspartyl from isoaspartyl products in peptides by electron capture dissociation. *Protein Sci.* 2005, *14*, 452–463.

[76] Cournoyer, J. J., Lin, C., O'Connor, P. B., Detecting deamidation products in proteins by electron capture dissociation. *Anal. Chem.* 2006, *78*, 1264–1271.

[77] O'Connor, P. B., Cournoyer, J. J., Pitteri, S. J., Chrisman, P. A., McLuckey, S. A., Differentiation of aspartic and isoaspartic acids using electron transfer dissociation. *J. Am. Soc. Mass Spectrom.* 2006, *17*, 15–19.

[78] Cournoyer, J. J., Lin, C., Bowman, M. J., O'Connor, P. B., Quantitating the relative abundance of isoaspartyl residues in deamidated proteins by electron capture dissociation. *J. Am. Soc. Mass Spectrom.* 2007, *18*, 48–56.

[79] Lehmann, W. D., Schlosser, A., Erben, G., Pipkorn, R. *et al.*, Analysis of isoaspartate in peptides by electrospray tandem mass spectrometry. *Protein Sci.* 2000, *9*, 2260–2268.

[80] Gonzalez, L. J., Shimizu, T., Satomi, Y., Betancourt, L. *et al.*, Differentiating alpha- and beta-aspartic acids by electrospray ionization and low-energy tandem mass spectrometry. *Rapid Commun. Mass. Spectrom.* 2000, *14*, 2092–2102.

[81] Pfeifer, N., Leinenbach, N., Huber, C. G., Kohlbacher, O., Improving peptide identification in proteome analysis by a two-dimensional retention time filtering approach. *J. Proteome Res.* 2009, *8*, 4109–4115.

[82] Spicer, V., Yamchuk, A., Cortens, J., Sousa, S. *et al.*, Sequence-specific retention calculator. A family of peptide retention time prediction algorithms in reversed-phase HPLC: Applicability to various chromatographic conditions and columns. *Anal. Chem.* 2007, *79*, 8762–8768.

[83] Krokhin, O. V., Craig, R., Spicer, V., Ens, W. *et al.*, An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC - Its application to protein peptide mapping by off-line HPLC-MALDI MS. *Mol. Cell. Proteomics* 2004, *3*, 908–919.

[84] Winter, D., Pipkorn, R., Lehmann, W. D., Separation of peptide isomers and conformers by ultra performance liquid chromatography. *J. Sep. Sci.* 2009, *32*, 1111–1119.

[85] Marshall, A. G., Hendrickson, C. L., High-resolution mass spectrometers. *Annu. Rev. Anal. Chem.* 2008, *1*, 579–599.

[86] Spengler, B., Accurate mass as a bioinformatic parameter in data-to-knowledge conversion: Fourier transform ion cyclotron resonance mass spectrometry for peptide de novo sequencing. *Eur. J. Mass Spectrom.* 2007, *13*, 83–87.

[87] Spengler, B., De novo sequencing, peptide composition analysis, and composition-based sequencing: A new strategy employing accurate mass determination by Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2004, *15*, 703–714.

[88] Frank, A. M., Savitski, M. M., Nielsen, M. L., Zubarev, R. A., Pevzner, P. A., De novo peptide sequencing and identification with precision mass spectrometry. *J. Proteome Res.* 2007, *6*, 114–123.

[89] Desiderio, D. M., Kai, M., Preparation of stable isotope-incorporated peptide internal standards for field desorption mass-spectrometry quantification of peptides in biological tissue. *Biomed. Mass Spectrom.* 1983, *10*, 471–479.

[90] Schnolzer, M., Jedrzejewski, P., Lehmann, W. D., Protease-catalyzed incorporation of O-18 into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* 1996, *17*, 945–953.

[91] Shevchenko, A., Chernushevich, I., Ens, W., Standing, K. G. *et al.*, Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass. Spectrom.* 1997, *11*, 1015–1024.

[92] Mo, W. J., Takao, T., Shimonishi, Y., Accurate peptide sequencing by post-source decay matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass. Spectrom.* 1997, *11*, 1829–1834.

[93] Uttenweiler-Joseph, S., Neubauer, G., Christoforidis, A., Zerial, M., Wilm, M., Automated de novo sequencing of proteins using the differential scanning technique. *Proteomics* 2001, *1*, 668–682.

[94] Bantscheff, M., Dumpelfeld, B., Kuster, B., Femtomol sensitivity post-digest O-18 labeling for relative quantification of differential protein complex composition. *Rapid Commun. Mass. Spectrom.* 2004, *18*, 869–876.

[95] Munchbach, M., Quadroni, M., Miotto, G., James, P., Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation directing moiety. *Anal. Chem.* 2000, *72*, 4047–4057.

[96] Cagney, G., Emili, A., De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nat. Biotechnol.* 2002, *20*, 163–170.

[97] Gu, S., Pan, S. Q., Bradbury, E. M., Chen, X., Use of deuterium-labeled lysine for efficient protein identification

and peptide de novo sequencing. *Anal. Chem.* 2002, *74*, 5774–5785.

[98] Hsu, J. L., Huang, S. Y., Chow, N. H., Chen, S. H., Stable-isotope dimethyl labeling for quantitative proteomics. *Anal. Chem.* 2003, *75*, 6843–6852.

[99] Noga, M. J., Asperger, A., Silberring, J., N-Terminal H-3/D-3-acetylation for improved high-throughput peptide sequencing by matrix-assisted laser desorption/ionization mass spectrometry with a time-of-flight/time-of-flight analyzer. *Rapid Commun. Mass. Spectrom.* 2006, *20*, 1823–1827.

[100] Lin, T., Glish, G. L., C-terminal peptide sequencing via multistage mass spectrometry. *Anal. Chem.* 1998, *70*, 5162–5165.

[101] Lin, T., Payne, A. H., Glish, G. L., Dissociation pathways of alkali-cationized peptides: Opportunities for C-terminal peptide sequencing. *J. Am. Soc. Mass Spectrom.* 2001, *12*, 497–504.

[102] Hager, J. W., A new linear ion trap mass spectrometer. *Rapid Commun. Mass. Spectrom.* 2002, *16*, 512–526.

[103] Hager, J. W., Le Blanc, J. C. Y., Product ion scanning using a Q-q-Q(linear ion trap) (Q TRAP (TM)) mass spectrometer. *Rapid Commun. Mass. Spectrom.* 2003, *17*, 1056–1064.

[104] Hager, J. W., Product ion spectral simplification using time-delayed fragment ion capture with tandem linear ion traps. *Rapid Commun. Mass. Spectrom.* 2003, *17*, 1389–1398.

[105] Lehmann, W. D., Single series peptide fragment ion spectra generated by two-stage collision-induced dissociation in a triple quadrupole. *J. Am. Soc. Mass Spectrom.* 1998, *9*, 606–611.

[106] Fernandez-de-Cossio, J., Gonzalez, J., Satomi, Y., Shima, T. *et al.*, Automated interpretation of low-energy collision-induced dissociation spectra by SeqMS, a software aid for de novo sequencing by tandem mass spectrometry. *Electrophoresis* 2000, *21*, 1694–1699.

[107] Liska, A. J., Shevchenko, A., Combining mass spectrometry with database interrogation strategies in proteomics. *Trac-Trends Anal. Chem.* 2003, *22*, 291–298.

[108] Mo, L. J., Dutta, D., Wan, Y. H., Chen, T., MSNovo: A dynamic programming algorithm for de novo peptide sequencing via tandem mass spectrometry. *Anal. Chem.* 2007, *79*, 4870–4878.

[109] Pevtsov, S., Fedulova, I., Mirzaei, H., Buck, C., Zhang, X., Performance evaluation of existing de novo sequencing algorithms. *J. Proteome Res.* 2006, *5*, 3018–3028.

[110] Pitzer, E., Masselot, A., Colinge, J., Assessing peptide de novo sequencing algorithms performance on large and diverse data sets. *Proteomics* 2007, *7*, 3051–3054.

[111] Savitski, M. M., Nielsen, M. L., Kjeldsen, F., Zubarev, R. A., Proteomics-grade de novo sequencing approach. *J. Proteome Res.* 2005, *4*, 2348–2354.

[112] Taylor, J. A., Johnson, R. S., Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry. *Anal. Chem.* 2001, *73*, 2594–2604.

[113] Shevchenko, A., Sunyaev, S., Loboda, A., Shevehenko, A. *et al.*, Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time of flight mass spectrometry and BLAST homology searching. *Anal. Chem.* 2001, *73*, 1917–1926.

[114] Junqueira, M., Spirin, V., Balbuena, T. S., Thomas, H. *et al.*, Protein identification pipeline for the homology-driven proteomics. *J. Proteomics* 2008, *71*, 346–356.

[115] Pearson, W. R., Lipman, D. J., Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 1988, *85*, 2444–2448.

[116] Shevchenko, A., Valcu, C. M., Junqueira, M., Tools for exploring the proteomosphere. *J. Proteomics* 2009, *72*, 137–144.

[117] Johnson, R. S., Biemann, K., The primary structure of thioredoxin from chromatium vinosum determined by high-performance tandem mass-spectrometry. *Biochemistry* 1987, *26*, 1209–1214.

[118] Branca, R. M. M., Bodo, G., Bagyinka, C., Prokai, L., De novo sequencing of a 21-kDa cytochrome c(4) from Thiocapsa roseopersicina by nanoelectrospray ionization ion-trap and Fourier-transform icon-cyclotron resonance mass spectrometry. *J. Mass Spectrom.* 2007, *42*, 1569–1582.

[119] Yague, J., Paradela, A., Ramos, M., Ogueta, S. *et al.*, Peptide rearrangement during quadrupole ion trap fragmentation: added complexity to MS/MS spectra. *Anal. Chem.* 2003, *75*, 1524–1535.

[120] Fodor, S., Zhang, Z. Q., Rearrangement of terminal amino acid residues in peptides by protease-catalyzed intramolecular transpeptidation. *Anal. Biochem.* 2006, *356*, 282–290.

[121] Schaefer, H., Chamrad, D. C., Marcus, K., Reidegeld, K. A. *et al.*, Tryptic transpeptidation products observed in proteome analysis by liquid chromatography-tandem mass spectrometry. *Proteomics* 2005, *5*, 846–852.

[122] Han, X. M., Jin, M., Breuker, K., McLafferty, F. W., Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science* 2006, *314*, 109–112.

[123] Yoo, C., Suckau, D., Sauerland, V., Ronk, M., Ma, M. H., Toward top-down determination of PEGylation site using MALDI in-source decay MS analysis. *J. Am. Soc. Mass Spectrom.* 2009, *20*, 326–333.

[124] Chait, B. T., Mass spectrometry: bottom-up or top-down? *Science* 2006, *314*, 65–66.

[125] Thompson, A. H., Bjourson, A. J., Orr, D. F., Shaw, C., McClean, S., A combined mass spectrometric and cDNA sequencing approach to the isolation and characterization of novel antimicrobial peptides from the skin secretions of *Phyllomedusa hypochondrialis* azurea. *Peptides* 2007, *28*, 1331–1343.

[126] Voss, C., Eyol, E., Frank, M., von der Lieth, C. W., Berger, M. R., Identification and characterization of riproximin, a new type II ribosome-inactivating protein with anti-neoplastic activity from *Ximenia americana*. *FASEB J.* 2006, *20*, E334–E345.

